

CERTAINTY BANDS FOR THE CONDITIONAL CUMULATIVE DISTRIBUTION FUNCTION AND APPLICATIONS

BY S. FERRIGNO¹, B. FOLIGUET², M. MAUMY-BERTRAND³, AND A. MULLER-GUEUDIN¹

ABSTRACT. In this paper, we establish uniform asymptotic certainty bands for the conditional cumulative distribution function. To this aim, we give exact rate of strong uniform consistency for the local linear estimator of this function. The corollaries of this result are the asymptotic certainty bands for the quantiles and the regression function. We illustrate our results with simulations and an application on fetopathologic data.

¹IECL (UMR 7502), Nancy-Université, CNRS, INRIA, members of the BIGS (BIology, Ge-
netics and Statistics) team at INRIA, France¹, ²Service de foetopathologie et de placentologie,
Maternité Régionale Universitaire, CHU Nancy, France², ³IRMA (UMR 7501), Université de
Strasbourg, France.³

Keywords: Conditional cumulative distribution function, local polynomial estimator, uniform asymptotic certainty bands, regression function, quantiles.

1. Introduction

1.1. Motivations. Consider (X, Y) , a random vector defined in $\mathbb{R} \times \mathbb{R}$. Here Y is the variable of interest and X the concomitant variable. Throughout, we work with a sample $\{(X_i, Y_i)_{1 \leq i \leq n}\}$ of independent and identically replica of (X, Y) . We will assume that (X, Y) [resp. X] has a density function $f_{X,Y}$ [resp. f_X] with respect to the Lebesgue measure. In this paper, we will mostly focus on the regression function of $\psi(Y)$ evaluate at $X = x$ defined by:

$$m_\psi(x) = \mathbb{E}(\psi(Y)|X = x) = \frac{1}{f_X(x)} \int_{\mathbb{R}} \psi(y) f_{X,Y}(x, y) dy, \quad \text{with } f_X(x) \neq 0 \quad (1)$$

whenever this regression function is meaningful. Here and elsewhere, ψ denotes a specified measurable function, which is assumed to be bounded on each compact subinterval of \mathbb{R} .

Because of numerous applications, the problem of estimating the function m_ψ , the density function f_X and the regression function $m_{\psi=\text{Id}}$ has been the subject of considerable interest during the last decades. We can cite for example Nadaraya [22], Watson [30], Devroye [10], Col-lomb [5], Härdle [18] and specially mention two articles, Einmahl and Mason [12] and Deheuvels and Mason [9] for two reasons. The first is that these articles study an estimator of m_ψ and its properties. The second is that we use the tools which are developed in these articles in order to establish our proofs. We now choose $\psi = \psi_t$ defined by $\psi_t(y) = \mathbb{1}_{\{y \leq t\}}$ with $t \in \mathbb{R}$ arbitrary but fixed, and $\mathbb{1}$ the indicator function, so we obtain the conditional cumulative distribution function (*cond-cdf*) of Y given $X = x$, defined by:

$$\forall t \in \mathbb{R}, \quad F(t|x) = m_{\psi_t}(x) = \mathbb{E}(\mathbb{1}_{\{Y \leq t\}}|X = x) = \mathbb{P}(Y \leq t|X = x). \quad (2)$$

¹sandie.ferrigno@univ-lorraine.fr, aurelie.gueudin@univ-lorraine.fr

²b.foliguete@maternite.chu-nancy.fr

³mmaumy@math.unistra.fr

Saying that, we are implicitly assuming the existence of a regular version for the conditional distribution of Y given X .

In this article, we study the conditional cumulative distribution function and a nonparametric estimator associated to this function. The *cond-cdf* has the advantages of completely characterizing the law of the random considered variable, allowing to obtain the regression function, the density function, the moments and the conditional quantile function. The *cond-cdf* is also used for example, in medicine (Gannoun *et al.* [16]) or econometric domain (Li *et al.* [20]).

Introduce the Nadaraya-Watson estimator (see Nadaraya [22] and Watson [30]) of the *cond-cdf* $F(t|x)$, for all $t \in \mathbb{R}$ and $x \in \mathbb{R}$, defined by:

$$\hat{F}_n^{(0)}(t, h_n|x) = \frac{\sum_{i=1}^n \mathbb{1}_{\{Y_i \leq t\}} K\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)} \text{ for } \sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right) \neq 0 \quad (3)$$

where $K(\cdot)$ is a positive-valued kernel function defined on \mathbb{R} and $(h_n)_{n \geq 1}$ is the bandwidth, and denotes a non-random sequence of positive constants satisfying some assumptions which will be defined latter.

For the study of the convergence rate of this estimator, it will be convenient to center $\hat{F}_n^{(0)}(t, h_n|x)$ by the estimator of $\mathbb{E}(\hat{F}_n^{(0)}(t, h_n|x))$:

$$\hat{\mathbb{E}}\left(\hat{F}_n^{(0)}(t, h_n|x)\right) = \frac{\mathbb{E}\left(\mathbb{1}_{\{Y \leq t\}} K\left(\frac{x-X}{h_n}\right)\right)}{\mathbb{E}\left(K\left(\frac{x-X}{h_n}\right)\right)}. \quad (4)$$

Remark 1.

- (1) With the hypothesis we have in mind (see in Section 1.2), the denominator of this quantity does not cancel.
- (2) In general, $\hat{\mathbb{E}}(\hat{F}_n^{(0)}(t, h_n|x))$ does not coincide with $\mathbb{E}(\hat{F}_n^{(0)}(t, h_n|x))$. However, under mild regularity assumptions, the difference between these two (non-random) quantities becomes asymptotically negligible as $h_n \searrow 0$ together with $nh_n \nearrow +\infty$ as $n \rightarrow +\infty$.

The estimator $\hat{F}_n^{(0)}(t, h_n|x)$ of the *cond-cdf* has been first treated by Collomb [4]. He proved consistency results, without rates, which are uniform in x and pointwise in t . A Glivenko-Cantelli type theorem for the Nadaraya-Watson estimator, uniform in t and pointwise in x , is given in Stute [27]. Moreover, Stute [26] was the first to obtain the exact rate of strong uniform consistency on compact intervals for the kernel density estimator. The best results for the other estimators in terms of approximate rates that we are aware of are due to Härdle *et al.* [19]. Such a result plays a fundamental role in obtaining strong uniform consistency rates in other statistics problems, like in the conditional quantile function.

More later, in 2000, Einmahl and Mason [12] have determined, under mild regularity conditions on the joint and marginal density functions and under hypotheses on the bandwidth (h_n) , exact rates of strong uniform consistency for the *cond-cdf*. We recall here their result:

Corollary 1. (see Corollary 2 in [12].) *Let I be a compact interval. Assume that $f_{X,Y}$ and f_X satisfy some regularity conditions and moreover that h_n satisfies $h_n \searrow 0$, $nh_n \nearrow +\infty$, $\log h_n^{-1} / \log \log n \rightarrow +\infty$ and $nh_n / \log n \rightarrow +\infty$ as $n \rightarrow +\infty$. Then we have for any*

kernel K defined in [12], with probability 1:

$$\lim_{n \rightarrow +\infty} \sup_{t \in \mathbb{R}} \sup_{x \in I} \sqrt{\frac{nh_n}{\log(h_n^{-1})}} \left| \widehat{F}_n^{(0)}(t, h_n|x) - \mathbb{E} \left(\widehat{F}_n^{(0)}(t, h_n|x) \right) \right| = \frac{\|K\|_2}{\sqrt{2 \inf_{x \in I} f_X(x)}} \quad (5)$$

where $\|K\|_2^2 = \int_{\mathbb{R}} K^2(u) du$.

Remark 2.

- (1) Under the assumptions of this corollary, the limit in Equation (5) does not depend on the distribution of the random variable Y .
- (2) In 2004, Blondin establishes in [3] a similar result of the Corollary 1 in the multivariate case, *i.e.* (X, Y) is in $\mathbb{R}^r \times \mathbb{R}^d$, $r, d \in \mathbb{N}^*$.

In 2005, Einmahl and Mason [13] have given an uniform in bandwidth consistency of kernel-type function estimators, in the case where (X, Y) is in $\mathbb{R}^r \times \mathbb{R}$, $r \in \mathbb{N}^*$, and specially for the estimator $\widehat{F}_n^{(0)}(t, h_n|x)$ defined in (3). We recall below their result:

Theorem 1.1. (see Theorem 3 in [13].) *Let I be a compact subset of \mathbb{R}^r and let K be a kernel defined in [13]. Suppose that f_X is continuous and strictly positive on J , which is a compact subset of \mathbb{R}^r and contains I . Then, with probability 1, we have for large enough $c > 0$ and any $b_n \searrow 0$:*

$$\limsup_{n \rightarrow +\infty} \sup_{c \log n \leq h \leq b_n} \sup_{t \in \mathbb{R}} \sup_{x \in I} \sqrt{\frac{nh}{\log(h^{-1}) \vee \log \log n}} \left| \widehat{F}_n^{(0)}(t, h|x) - \mathbb{E} \left(\widehat{F}_n^{(0)}(t, h|x) \right) \right| < +\infty. \quad (6)$$

Remark 3. In this result, the exact value of the limit is unknown.

It is a well-known fact the asymptotic bias of the Nadaraya-Watson estimator has a bad form. To overcome this problem, there exists an alternative: the local polynomial techniques described in Fan and Gijbels [14] or in Tsybakov [29].

To study the local polynomial estimators, either we can use the U -statistics, for example Mint El Mouvid [21]. But this method implies heavy calculations. Or we can use the empirical processes, for example Dony *et al.* [11]. But the results on the empirical processes indexed by classes of functions are established only for classes of real-valued functions.

The present paper is organized as follows. First, we introduce the local linear estimator of the *cond-cdf*, with the main notations and assumptions needed for our task. Then we establish an uniform law of the logarithm for the local linear estimator of the *cond-cdf* in Section 2. In Section 3, we show that limit laws of the logarithm are useful in the construction of uniform asymptotic certainty bands for the *cond-cdf*, the regression function and the conditional quantile function. Such certainty bands are obtained from simulations in Section 4 and from fetopathologic data in Section 5. Finally, Section 6 is devoted to the proofs of our results.

1.2. Notations and assumptions. Let $(X_1, Y_1), (X_2, Y_2), \dots$, be independent and identically distributed replica of (X, Y) in $\mathbb{R} \times \mathbb{R}$. Let $I = [a, b], J = [a', b'] \supseteq I$, two fixed compacts of \mathbb{R} .

First, we impose the following set of assumptions upon the distribution of (X, Y) :

- (F.1) $f_{X,Y}$ is continuous on $J \times \mathbb{R}$ and f_X is continuous and strictly positive on J ;
- (F.2) $Y \mathbb{1}_{\{X \in J\}}$ is bounded on \mathbb{R} .

Remark 4.

- (1) Under (F.1-2), the *cond-cdf* is well defined.

- (2) The assumption (F.2) is very useful for the proof of our results. This boundedness assumption is common in non-parametric estimation. It ensures the existence of several moments of the *cond-cdf*.

K denotes a positive-valued kernel function defined on \mathbb{R} , fulfilling the conditions:

- (K.1) K is right-continuous function with bounded variation on \mathbb{R} ;
- (K.2) K is compactly supported and $\int_{\mathbb{R}} K(u)du = 1$;
- (K.3) $\int_{\mathbb{R}} uK(u)du = 0$ and $\int_{\mathbb{R}} u^2K(u)du \neq 0$.

We note: $\|K\|_2^2 = \int_{\mathbb{R}} K^2(u)du$.

Further, introduce the following assumptions on the non-random sequence $(h_n)_{n \geq 1}$:

- (H.0) for all n , $0 < h_n < 1$;
- (H.1) $h_n \rightarrow 0$, as $n \rightarrow +\infty$;
- (H.2) $nh_n/\log n \rightarrow +\infty$, as $n \rightarrow +\infty$;
- (H.3) $h_n \searrow 0$ and $nh_n \nearrow +\infty$, as $n \rightarrow +\infty$;
- (H.4) $\log(h_n^{-1})/\log \log n \rightarrow +\infty$, as $n \rightarrow +\infty$.

Remark 5.

- (1) The assumption (H.0) is necessary to define $\sqrt{\log(h_n^{-1})}^{-1}$ (see later in our Theorem 2.1).
- (2) The assumptions (H.0-2) are necessary and sufficient for our uniform convergence in probability (see Theorem 2.1).
- (3) In order to have almost surely convergence results, we need the assumptions (H.3-4) (see Blondin [3]).
- (4) The assumptions (H.0, H.2-4) are called the Csörgö-Révész-Stute assumptions.

Our aim will be to establish the strong uniform consistency of the local linear estimator of the conditional cumulative distribution function, defined by:

$$\widehat{F}_n^{(1)}(t, h_n|x) = \frac{\widehat{f}_{n,2}(x, h_n)\widehat{r}_{n,0}(x, t, h_n) - \widehat{f}_{n,1}(x, h_n)\widehat{r}_{n,1}(x, t, h_n)}{\widehat{f}_{n,0}(x, h_n)\widehat{f}_{n,2}(x, h_n) - \left(\widehat{f}_{n,1}(x, h_n)\right)^2} \quad (7)$$

where $^{(1)}$ denotes the order 1 of the local polynomial estimator, and

$$\widehat{f}_{n,j}(x, h_n) = \frac{1}{nh_n} \sum_{i=1}^n \left(\frac{x - X_i}{h_n} \right)^j K \left(\frac{x - X_i}{h_n} \right), \text{ for } j = 0, 1, 2, \quad (8)$$

$$\widehat{r}_{n,j}(x, t, h_n) = \frac{1}{nh_n} \sum_{i=1}^n \mathbb{1}_{\{Y_i \leq t\}} \left(\frac{x - X_i}{h_n} \right)^j K \left(\frac{x - X_i}{h_n} \right), \text{ for } j = 0, 1. \quad (9)$$

Remark 6.

- (1) The Nadaraya-Watson estimator $\widehat{F}_n^{(0)}(t, h_n|x)$ can be also written with the functions $\widehat{f}_{n,j}$ and $\widehat{r}_{n,j}$ as

$$\widehat{F}_n^{(0)}(t, h_n|x) = \frac{\widehat{r}_{n,0}(x, t, h_n)}{\widehat{f}_{n,0}(x, h_n)}.$$

It is the local polynomial estimator of order 0 of the conditional cumulative distribution function.

- (2) The estimator $\widehat{F}_n^{(1)}(t, h_n|x)$ is better than the Nadaraya-Watson estimator when the design is random and has the favorable property to reproduce polynomial of order 1. Precisely, the local linear estimator has a high minimax efficiency among all possible estimators, including nonlinear smoothers (see Fan and Gijbels [14]).
- (3) We have state in the beginning of this Section that we restrict ourselves to the local polynomial estimator of order 1. The local polynomial estimator can be generalized to the orders $p \geq 2$, but the equations become more complicated. We show briefly the form of the local polynomial estimator of order 2:

$$\widehat{F}_n^{(2)}(t, h_n|x) = \frac{a_1 \widehat{r}_{n,0}(x, t, h_n) + a_2 \widehat{r}_{n,1}(x, t, h_n) + a_3 \widehat{r}_{n,2}(x, t, h_n)}{a_1 \widehat{f}_{n,0}(x, h_n) + a_2 \widehat{f}_{n,1}(x, h_n) + a_3 \widehat{f}_{n,2}(x, h_n)}$$

where $\begin{cases} a_1 = \widehat{f}_{n,2}(x, h_n) \widehat{f}_{n,4}(x, h_n) - \left(\widehat{f}_{n,3}(x, h_n) \right)^2 \\ a_2 = \widehat{f}_{n,2}(x, h_n) \widehat{f}_{n,3}(x, h_n) - \widehat{f}_{n,1}(x, h_n) \widehat{f}_{n,4}(x, h_n) \\ a_3 = \widehat{f}_{n,1}(x, h_n) \widehat{f}_{n,3}(x, h_n) - \left(\widehat{f}_{n,2}(x, h_n) \right)^2 \end{cases}$

and $\widehat{f}_{n,3}$, $\widehat{f}_{n,4}$ and $\widehat{r}_{n,2}$ are the direct extensions of the definitions given in the Equations (8) and (9). Note also that, it is not very interesting to study $p \geq 3$, see Fan and Gijbels [14], pp. 20-22 and 77-80. The argument is that the mean square error increases with p .

Now, we study the consistency of the estimator $\widehat{F}_n^{(1)}(t, h_n|x)$ via the following decomposition:

$$\widehat{F}_n^{(1)}(t, h_n|x) - F(t|x) = \underbrace{\widehat{F}_n^{(1)}(t, h_n|x) - \widehat{\mathbb{E}} \left(\widehat{F}_n^{(1)}(t, h_n|x) \right)}_{(1)} + \underbrace{\widehat{\mathbb{E}} \left(\widehat{F}_n^{(1)}(t, h_n|x) \right) - F(t|x)}_{(2)}$$

where, following the ideas of Deheuvels and Mason (see [9]), the centering term is defined by:

$$\widehat{\mathbb{E}} \left(\widehat{F}_n^{(1)}(t, h_n|x) \right) = \frac{f_{n,2}(x, h_n) r_{n,0}(x, t, h_n) - f_{n,1}(x, h_n) r_{n,1}(x, t, h_n)}{f_{n,0}(x, h_n) f_{n,2}(x, h_n) - f_{n,1}^2(x, h_n)}$$

where $f_{n,j}(x, h_n) = \mathbb{E} \left(\widehat{f}_{n,j}(x, h_n) \right)$ for $j = 0, 1, 2$ and $r_{n,j}(x, t, h_n) = \mathbb{E} \left(\widehat{r}_{n,j}(x, h_n) \right)$ for $j = 0, 1$.

The *random part* (1) is the object of our theorem given in the following Section. Under (F.1-2), (H.1) and (K.1-3), the *deterministic term* (2), so-called bias, converges uniformly to 0 over $(x, t) \in I \times \mathbb{R}$. The argument to proof this is the Bochner's Lemma (see for instance [12], or our Equations (24) in Section 6).

2. Uniform consistency of the local linear estimator

We have now all the ingredients to state our main results. The uniform law of the logarithm concerning the local linear estimator of the *cond-cdf*, is given in Theorem 2.1 below.

Theorem 2.1. *Under (F.1-2), (H.0-2) and (K.1-3), we have:*

$$\sup_{x \in I} \sqrt{\frac{nh_n}{\log(h_n^{-1})}} \left| \widehat{F}_n^{(1)}(t, h_n|x) - \widehat{\mathbb{E}} \left(\widehat{F}_n^{(1)}(t, h_n|x) \right) \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \sigma_{F,t}(I) \quad (10)$$

where $\sigma_{F,t}^2(I) = 2 \|K\|_2^2 \sup_{x \in I} \left(\frac{F(t|x)(1-F(t|x))}{f_X(x)} \right)$.

Moreover, we have:

$$\sup_{t \in \mathbb{R}} \sup_{x \in I} \sqrt{\frac{nh_n}{\log(h_n^{-1})}} \left| \widehat{F}_n^{(1)}(t, h_n|x) - \widehat{\mathbb{E}} \left(\widehat{F}_n^{(1)}(t, h_n|x) \right) \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \sigma_F(I) \quad (11)$$

where

$$\sigma_F^2(I) = 2\|K\|_2^2 \sup_{t \in \mathbb{R}} \sup_{x \in I} \left(\frac{F(t|x)(1 - F(t|x))}{f_X(x)} \right) = \frac{\|K\|_2^2}{2 \inf_{x \in I} f_X(x)}.$$

The proof of Theorem 2.1 is postponed to Section 6.

Remark 7.

- (1) The matching almost surely convergence result can also be obtained by assuming (H.2-4) instead of (H.0-2).
- (2) The terms $\sigma_{F,t}(I)$ and $\sigma_F(I)$ depend upon the unknown density f_X . But it is a minor problem in practice, because, as shown in Deheuvels [7], and Deheuvels and Mason [9], an application of Slutsky's Lemma allows us to replace, without loss of generality, this quantity by $\widehat{f}_{n,0}(x, h_n)$ (or by any other estimator of $f_X(x)$ which is uniformly consistent on I). Indeed, under (F.1-2), (H.0-2), (K.1-3) we have $\sup_{x \in I} \left| \frac{\widehat{f}_{n,0}(x, h_n)}{f_X(x)} - 1 \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0$.

This last remark yields to the following corollary.

Corollary 2. *Under (F.1-2), (H.0-2), (K.1-3), we have:*

$$\sup_{t \in \mathbb{R}} \sup_{x \in I} \sqrt{\frac{2nh_n}{\|K\|_2^2 \log(h_n^{-1})}} \widehat{f}_{n,0}(x, h_n) \left| \widehat{F}_n^{(1)}(t, h_n|x) - \widehat{\mathbb{E}} \left(\widehat{F}_n^{(1)}(t, h_n|x) \right) \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 1. \quad (12)$$

We introduce the following quantity $L_n(x) := \sqrt{\frac{2nh_n}{\|K\|_2^2 \log(h_n^{-1})}} \widehat{f}_{n,0}(x, h_n)^{-1}$. We have noted at the end of the Section 1 that the bias part can be neglected, then we have the following proposition.

Proposition 1. *Under (F.1-2), (H.0-2) and (K.1-3), and if h_n is such that the bias term $\sup_{t \in \mathbb{R}} \sup_{x \in I} \{L_n(x)\}^{-1} \left| F(t|x) - \widehat{\mathbb{E}} \left(\widehat{F}_n^{(1)}(t, h_n|x) \right) \right| \xrightarrow[n \rightarrow +\infty]{} 0$ then we have:*

$$\sup_{t \in \mathbb{R}} \sup_{x \in I} \{L_n(x)\}^{-1} \left| \widehat{F}_n^{(1)}(t, h_n|x) - F(t|x) \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 1. \quad (13)$$

Remark 8.

- (1) The matching almost surely convergence result can also be obtained by assuming (H.2-4) instead of (H.0-2).
- (2) For our applications in Sections 4 and 5, a reference choice for h_n is given by minimizing the weighted Mean Integrated Square Error (MISE) criteria (see for instance Berlinet [2], Deheuvels [6] or Deheuvels and Mason [9]). A detailed discussion about the theoretical choice of this bandwidth is given in Ferrigno [15]. The asymptotically optimal constant bandwidth is given by:

$$h_n = C(K, F, f_X) n^{-\frac{1}{5}}$$

where the constant $C(K, F, f_X)$ is easy to calculate.

- (3) The choice of the kernel K is not important in practice. The most common used kernels are the Gaussian, the indicator function over $[-\frac{1}{2}, \frac{1}{2}]$, and the Epanechnikov kernels (see for instance Deheuvels [6]). Note that the Gaussian kernel is not compactly supported, but our results can be extended to this case.

3. Uniform asymptotic certainty bands

3.1. Application to the cond-cdf. We show now how the Proposition 1 can be used to construct uniform asymptotic certainty bands for $F(t|x)$, in the following sense. Under the assumptions of the Proposition 1, we have, for each $0 < \varepsilon < 1$, and as $n \rightarrow +\infty$:

$$\mathbb{P} \left\{ F(t|x) \in \left[\widehat{F}_n^{(1)}(t, h_n|x) \pm (1 + \varepsilon)L_n(x) \right], \text{ for all } (x, t) \in I \times \mathbb{R} \right\} \rightarrow 1 \quad (14)$$

and

$$\mathbb{P} \left\{ F(t|x) \in \left[\widehat{F}_n^{(1)}(t, h_n|x) \pm (1 - \varepsilon)L_n(x) \right], \text{ for all } (x, t) \in I \times \mathbb{R} \right\} \rightarrow 0. \quad (15)$$

Whenever (14) and (15) hold jointly for each $0 < \varepsilon < 1$, we have the following corollary:

Corollary 3. *Under (F.1-2), (H.0-2) and (K.1-3), and if h_n is such that the bias term $\sup_{t \in \mathbb{R}} \sup_{x \in I} \{L_n(x)\}^{-1} |F(t, h_n|x) - \widehat{\mathbb{E}}(\widehat{F}_n^{(1)}(t, h_n|x))| \xrightarrow{n \rightarrow +\infty} 0$ then the interval*

$$\left[\widehat{F}_n^{(1)}(t, h_n|x) \pm L_n(x) \right] \quad (16)$$

provides uniform asymptotic certainty bands (at an asymptotic confidence level of 100%) for the cond-cdf $F(t|x)$, uniformly in $(x, t) \in I \times \mathbb{R}$.

Remark 9.

- (1) Probability convergence is sufficient for forming certainty bands, and requires less restrictive hypotheses on the bandwidth h_n than the almost surely convergence results. That is why we use only the probability convergence result of the Proposition 1.
- (2) Following a suggestion of Deheuvels and Derzko [8], we use, for these upper and lower bounds for $F(t|x)$, the qualification of certainty bands, rather than of confidence bands, because there is no preassigned confidence level $\alpha \in (0, 1)$. Some authors (see for instance Deheuvels and Mason [9], or Blondin [3]) have used the term confidence bands.

3.2. Application to the regression function. Let $m(x) = \mathbb{E}(Y|X = x)$ the regression function and $\widehat{m}_n^{(1)}(x) = \int y \widehat{F}_n^{(1)}(dy, h_n|x)$ its local linear estimator. The Proposition 1 has the following corollary for the regression function.

Corollary 4. *Under (F.1-2), (H.0-2) and (K.1-3), and if h_n is such that the bias term $\sup_{t \in \mathbb{R}} \sup_{x \in I} \{L_n(x)\}^{-1} \left| F(t|x) - \widehat{\mathbb{E}}(\widehat{F}_n^{(1)}(t, h_n|x)) \right| \xrightarrow{n \rightarrow +\infty} 0$ and the variable Y lives in the real interval $[\alpha, \beta]$, then the interval*

$$\left[\widehat{m}_n^{(1)}(x) \pm (\beta - \alpha)L_n(x) \right] \quad (17)$$

provides uniform asymptotic certainty bands (at an asymptotic confidence level of 100%) for the conditional regression function $m(x)$, uniformly in $x \in I$.

The proof of Corollary 4 is postponed to Section 6.

3.3. Application to the conditional quantiles. Let $0 < \alpha < 1$. We define the conditional α -quantile of the cond-cdf by:

$$q_\alpha(x) = \inf\{t \in \mathbb{R} : F(t|x) \geq \alpha\}, \quad \text{for all } \alpha \in (0, 1).$$

The local linear estimator of the conditional α -quantile is defined by:

$$\widehat{q}_{\alpha,n}^{(1)}(x) = \inf\{t \in \mathbb{R} : \widehat{F}_n^{(1)}(t, h_n|x) \geq \alpha\}, \quad \text{for all } \alpha \in (0, 1).$$

The Proposition 1 has the following corollary for the conditional quantiles.

Corollary 5. *Under (F.1-2), (H.0-2) and (K.1-3), if h_n is such that the bias term $\sup_{t \in \mathbb{R}} \sup_{x \in I} \{L_n(x)\}^{-1} \left| F(t|x) - \mathbb{E} \left(\hat{F}_n^{(1)}(t, h_n|x) \right) \right| \xrightarrow{n \rightarrow +\infty} 0$ and if the function $x \mapsto f_{X,Y}(x, q_\alpha(x)) \neq 0$ for all $x \in I$, then the interval*

$$\left[\hat{q}_{\alpha,n}^{(1)}(x) \pm \frac{2L_n(x)f_X(x)}{f_{X,Y}(x, q_\alpha(x))} \right] \quad (18)$$

provides uniform asymptotic certainty bands (at an asymptotic confidence level of 100%) for the conditional α -quantile $q_\alpha(x)$, uniformly in $x \in I$.

The proof of Corollary 5 is postponed to Section 6.

Remark 10.

- (1) The form of these certainty bands is not very useful in practice since the bounds depend upon the unknown conditional density $f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$. Nevertheless, this gives the order of the deviation $|\hat{q}_{\alpha,n}^{(1)}(x) - q_\alpha(x)|$.
- (2) To give a more practical result, the idea is to replace the conditional density $f_{Y|X}(q_\alpha(x)|x)$ by an estimator $\hat{f}_{Y|X}(\hat{q}_{\alpha,n}^{(1)}(x)|x)$ such that $\sup_{x \in I} \left| \frac{\hat{f}_{Y|X}(\hat{q}_{\alpha,n}^{(1)}(x)|x)}{f_{Y|X}(q_\alpha(x)|x)} - 1 \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0$. This is not the object of the present article, and will be presented in a future work. A review of kernel estimators for the conditional density is given for instance in [32, 31]. We can cite here the kernel estimator of Parzen-Rosenblatt [23, 24].

4. A simulation study

In this paragraph, the *cond-cdf* and the certainty bands introduced in Corollary 3 are constructed on simulated data. We considered the case: $X \sim \mathcal{N}(0, 1)$ where $\mathcal{N}(0, 1)$ denotes the Gaussian distribution with mean 0 and standard deviation 1. We present two models:

- (M₁) $Y|X = x$ follows a Beta(a, b) distribution with shape parameters $a = 1$ and $b = 1 + x^2$.
- (M₂) $Y|X = x$ follows an Uniform distribution between $-|x|$ and $|x|$.

We worked with the sample sizes $n = 100$ and $n = 500$. For the kernel K , we opted for the Epanechnikov kernel. For the bandwidth, we selected $h_n = n^{-1/5}$. The Figure 1 illustrates the results for the models (M₁) and (M₂) defined above. For each model, we give the graph of a sample $(X_i, Y_i)_{i=1, \dots, n}$, and the *cond-cdf*: the true function $F(t|x)$ is in full line, whereas the estimated conditional distribution $\hat{F}_n^{(1)}(t, h_n|x)$ is in black dashed line, and certainty bands in grey line, for $x = 0$ and 1.

The confidence bands appear to be adequate. The fact that the true function does not belong to our certainty bands for some points was expected: it is due to the ε term in Equations (14) and (15). For $n = 500$, the results are better than for $n = 100$.

5. Application in study of the fetal growth

The study is based on 3606 fetuses autopsied in fetopathologic units of the "Service de foetopathologie et de placentologie" of the Maternité Régionale Universitaire (CHU Nancy, France) between 1996 and 2013. From this dataset, 694 fetuses were carefully selected by exclusion of multiple pregnancies, malformed, macerated or serious ill fetuses, or those with chromosomal abnormalities.

The naive idea, classically used by the fetopathologists or the echographists (see for instance [1], [25]), is to fit a parametric regression model $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \epsilon_i$ with the assumptions that ϵ_i , for $i = 1, \dots, n$ are independent and follow the Gaussian distribution

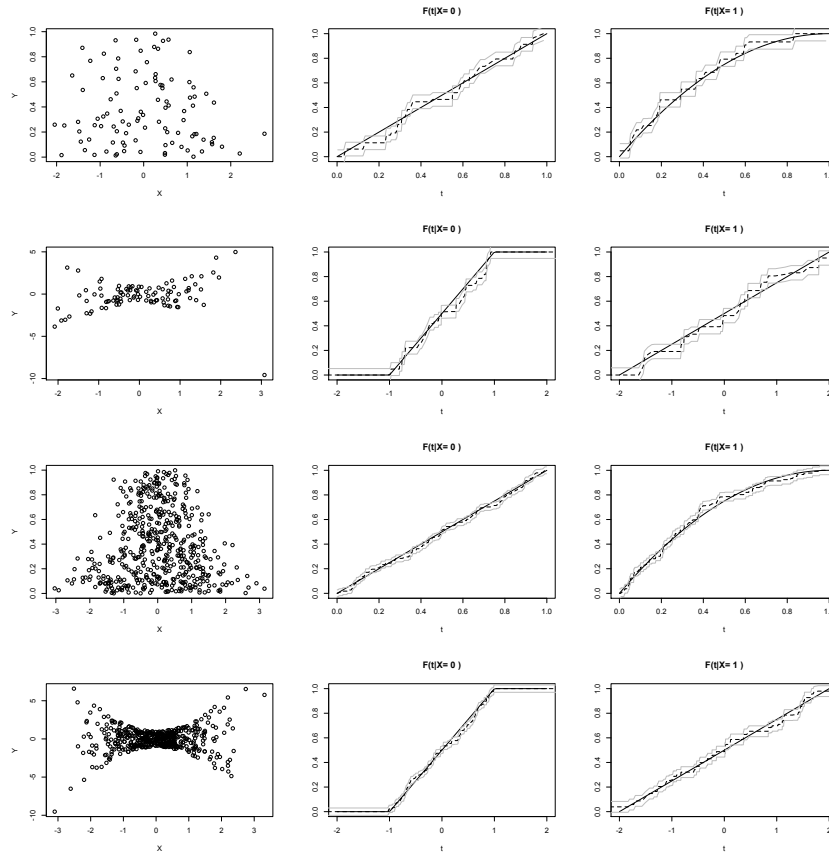


FIGURE 1. From top to bottom: models (M_1) and (M_2) for $n = 100$, and (M_1) and (M_2) for $n = 500$.

$\mathcal{N}(0, \sigma)$. The parameters $\beta_0, \beta_1, \beta_2, \sigma$ are estimated by the least squares method. We use the R 2.15.1 function `lm`.

The result is shown on the left graph of the Figure 2. This method yields to several problems:

- We obtain heteroscedastic and non-Gaussian errors.
- Moreover, regarding the confidence intervals of the previsions, they show that the prevision uncertainty is not growing with the gestational week: this is not consistent with the medical intuition.
- Another problem is that the global polynomial estimation can not enhance some changes in the growing curve of the fetal weight. For the fetopathologists, such changes are important as they correspond to delicate periods during the intrauterine growth. These change points can not be observed by a global estimation.

For these reasons, the local polynomial estimation is then a non-parametric alternative to the global parametric regression model.

We can conclude, by the observation of the right graph of the Figure 2:

- Our method gives the mean, the confidence intervals and the median weight. Satisfactorily, the confidence intervals show the growing of the prevision uncertainty with the gestational week.
- We observe for instance a change point between the 20th and 25th gestational week on the 0.975 percentile curve. This change point corresponds to the viability date of

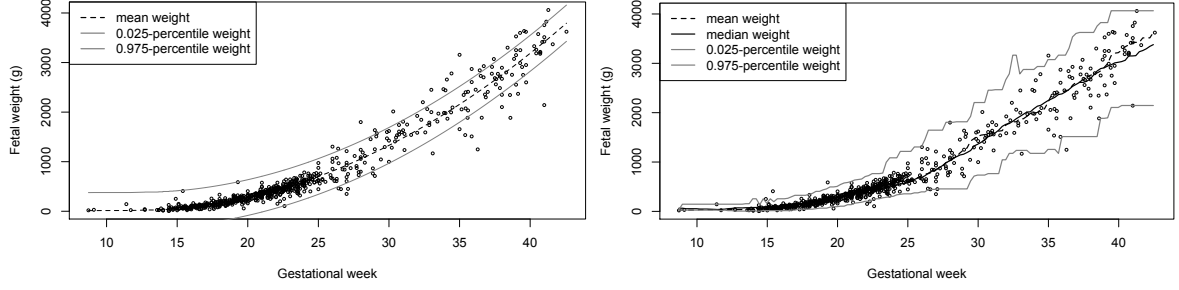


FIGURE 2. Fetal weight during the pregnancy: estimation of mean and quantiles with the second order polynomial regression (left), and with the linear local method (right).

the fetus. We can also remark a decrease of the growing speed around the 35th week. This has also been remarked in the medical article [17], where it is explained that this time corresponds to the regression (in the medical sense) of the placenta. More precise statistical tests to detect the change points of the fetal growth will be presented in a future work.

6. Proofs

6.1. Proof of Theorem 2.1. We prove the probability convergence result (Theorem 2.1), uniformly in $(x, t) \in I \times \mathbb{R}$. The uniform in $x \in I$ result (less difficult) is left to the reader.

Remark 11. The almost surely convergence result could be proved with some additional arguments, based on the Borel-Cantelli's Lemma (see for instance Blondin [3] or Einmahl and Mason [12]).

Step 1: In a first step, we introduce a general local empirical process. For any $j = 0, 1, 2$ and continuous real valued functions $c(\cdot)$ and $d(\cdot)$ on J , set for $x \in J$, $t \in \mathbb{R}$,

$$W_{n,j}(x, t) = \sum_{i=1}^n (c(x)\mathbb{1}_{\{Y_i \leq t\}} + d(x)) K_j \left(\frac{x - X_i}{h_n} \right) - n\mathbb{E} \left((c(x)\mathbb{1}_{\{Y \leq t\}} + d(x)) K_j \left(\frac{x - X}{h_n} \right) \right) \quad (19)$$

where

$$K_j(u) = u^j K(u) \quad (20)$$

for $j = 0, 1, 2$ and $u \in \mathbb{R}$.

For every fixed $t \in \mathbb{R}$, and $j = 0, 1, 2$, the process $W_{n,j}(\cdot, t)$ can be represented as a bivariate empirical process indexed by a class of functions. More precisely, we have:

$$W_{n,j}(\cdot, t) = \sqrt{n}\alpha_n(g) = \sum_{i=1}^n \{g(X_i, Y_i) - \mathbb{E}(g(X, Y))\},$$

where α_n is the empirical process based upon $(X_1, Y_1), \dots, (X_n, Y_n)$ and indexed by a suitable subclass $\mathcal{F}_{n,j}$ of the class of functions defined on $J \times \mathbb{R}$:

$$\mathcal{F}_j = \left\{ (x, y) \mapsto \{c(z)\mathbb{1}_{\{y \leq t\}} + d(z)\} K_j \left(\frac{z - x}{h} \right) : t \in \mathbb{R}, z \in I, 0 < h < 1 \right\}.$$

We give now the result from which our main Theorem 2.1 follows.

Theorem 6.1. *Under (F.1-2), (H.0-2), (K.1-3) we have:*

$$\sqrt{\frac{1}{2nh_n \log(h_n^{-1})}} \sup_{t \in \mathbb{R}} \sup_{x \in I} |W_{n,j}(x, t)| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \sigma_{W,j}(I), \quad (21)$$

where

$$\sigma_{W,j}^2(I) = \sup_{t \in \mathbb{R}} \sup_{x \in I} \mathbb{E} \left(|c(x) \mathbb{1}_{\{Y \leq t\}} + d(x)|^2 |X = x \right) f_X(x) \int_{\mathbb{R}} K_j^2(u) du.$$

Proof: The proof is divided into an upper bound result, and a lower bound result.

Upper bound part: The proof of the upper bound result is divided into two steps. The hypothesis (F.2) is important in the upper bound part: we can find a real $M > 0$ such that if $X \in J$ then $|Y| \leq M$.

Step A: Discretization in $x \in I$ and $t \in [-M, M]$. First, we examine the behavior of our process $(x, t) \mapsto W_{n,j}(x, t)$ on an appropriate chosen grid of $I \times [-M, M]$, with increment δh_n for I and increment δ for $[-M, M]$, for fixed $0 < \delta < 1$:

$$\begin{cases} z_{i,n} = a + i\delta h_n, & i = 1, \dots, i_n = \left\lfloor \frac{b-a}{\delta h_n} \right\rfloor, \\ t_l = -M + l\delta, & l = 1, \dots, L = \left\lfloor \frac{2M}{\delta} \right\rfloor, \end{cases}$$

where $[u] \leq u < [u] + 1$ represents the integer part of u .

The study of the supremum on $I \times [-M, M]$ is then reduced to the study of the maximum on a finite number of points. The empirical process is then indexed on the finite class of functions

$$\mathcal{F}_{n,j} = \left\{ (x, y) \mapsto \{c(z_{i,n}) \mathbb{1}_{y \leq t_l} + d(z_{i,n})\} K_j \left(\frac{z_{i,n} - x}{h_n} \right) : i = 1, \dots, i_n, l = 1, \dots, L \right\}.$$

The useful tool in this Step is the Bernstein inequality (see for instance Deheuvels and Mason [9]).

Step B: Oscillation. Next we study the behavior of our process between the grid points $z_{i,n}$ and t_l for $1 \leq i \leq i_n$ and $1 \leq l \leq L$. The objective of this Step is to study the maximal oscillations between the grid points. The useful tool in this Step is the Talagrand inequality for VC Classes (see for instance Talagrand [28], Einmahl and Mason [12] or Blondin [3]).

Lower bound part: The lower bound result is proved with technical results based on Poisson processes, and needs the Bochner's Lemma. In this part, the hypothesis (F.1) is particularly important.

Step 2: We give now useful corollaries of Theorem 6.1. We recall that K_j has been defined in Equation (20), and $\|K_j\|_2^2 = \int_{\mathbb{R}} K_j^2(u) du$.

Corollary 6. *Under (F.1-2), (H.0-2) and (K.1-3), we have, by application of Theorem 6.1 with $c(x) = 0, d(x) = 1, j = 0, 1, 2$:*

$$\sqrt{\frac{nh_n}{2 \log(h_n^{-1})}} \sup_{x \in I} |\hat{f}_{n,j}(x) - f_{n,j}(x)| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \sigma_{f,j}(I), \quad (22)$$

where

$$\sigma_{f,j}^2(I) = \|K_j\|_2^2 \sup_{x \in I} \{f_X(x)\}.$$

Corollary 7. *Under (F.1-2), (H.0-2) and (K.1-3), we have, by application of Theorem 6.1 with $c(x) = 1, d(x) = 0, j = 0, 1$:*

$$\sqrt{\frac{nh_n}{2\log(h_n^{-1})}} \sup_{t \in \mathbb{R}} \sup_{x \in I} |\hat{r}_{n,j}(x, t) - r_{n,j}(x, t)| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \sigma_{r,j}(I) \quad (23)$$

where

$$\sigma_{r,j}^2(I) = \|K_j\|_2^2 \sup_{t \in \mathbb{R}} \sup_{x \in I} \{F(t|x) f_X(x)\}.$$

Moreover, under (F.1-2), (H.1) and (K.1-3), the Bochner's Lemma (cf for instance [12]) implies, uniformly in $(x, t) \in I \times \mathbb{R}$:

$$\begin{aligned} f_{n,0}(x) &= f_X(x) \mu_0(K) + o(1), \\ f_{n,1}(x) &= f_X(x) \mu_1(K) + o(1), \\ f_{n,2}(x) &= f_X(x) \mu_2(K) + o(1), \\ r_{n,0}(x, t) &= f_X(x) F(t|x) + o(1), \\ r_{n,1}(x, t) &= f_X(x) F(t|x) \mu_1(K) + o(1), \end{aligned} \quad (24)$$

where $\mu_j(K) = \int_{\mathbb{R}} K_j(u) du$, for $j = 0, 1, 2$. The hypotheses (K.2-3) imply that $\mu_0(K) = 1$, $\mu_1(K) = 0$ and $\mu_2(K) \neq 0$.

Step 3: In this third step, the deviation $\hat{F}_n^{(1)}(t|x) - \mathbb{E}(\hat{F}_n^{(1)}(t|x))$ can be asymptotically expressed as a linear function of the bivariate empirical process.

$$\begin{aligned} &\hat{F}_n^{(1)}(t|x) - \mathbb{E}(\hat{F}_n^{(1)}(t|x)) \\ &= \frac{\hat{r}_{n,0}(x, t) \hat{f}_{n,2}(x) - \hat{f}_{n,1}(x) \hat{r}_{n,1}(x, t)}{\hat{f}_{n,2}(x) \hat{f}_{n,0}(x) - \hat{f}_{n,1}^2(x)} - \frac{r_{n,0}(x, t) f_{n,2}(x) - f_{n,1}(x) r_{n,1}(x, t)}{f_{n,2}(x) f_{n,0}(x) - f_{n,1}^2(x)} \quad (25) \\ &= \frac{\hat{r}_{n,0}(x, t) \hat{f}_{n,2}(x) - r_{n,0}(x, t) f_{n,2}(x) + f_{n,1}(x) r_{n,1}(x, t) - \hat{f}_{n,1}(x) \hat{r}_{n,1}(x, t)}{\hat{f}_{n,2}(x) \hat{f}_{n,0}(x) - \hat{f}_{n,1}^2(x)} \quad (26) \\ &+ \frac{(r_{n,0}(x, t) f_{n,2}(x) - f_{n,1}(x) r_{n,1}(x, t))(f_{n,2}(x) f_{n,0}(x) - f_{n,1}^2(x) - \hat{f}_{n,2}(x) \hat{f}_{n,0}(x) + \hat{f}_{n,1}^2(x))}{(\hat{f}_{n,2}(x) \hat{f}_{n,0}(x) - \hat{f}_{n,1}^2(x))(f_{n,2}(x) f_{n,0}(x) - f_{n,1}^2(x))} \quad (27) \end{aligned}$$

First, studying the numerator of the expression (26), we have:

$$\begin{aligned} \text{Num}(26) &= \left((\hat{f}_{n,2}(x) - f_{n,2}(x)) + (f_{n,2}(x) - f_X(x) \mu_2(K)) + f_X(x) \mu_2(K) \right) (\hat{r}_{n,0}(x, t) - r_{n,0}(x, t)) \\ &+ \left((r_{n,0}(x, t) - f_X(x) F(t|x)) + f_X(x) F(t|x) \right) (\hat{f}_{n,2}(x) - f_{n,2}(x)) \\ &+ f_{n,1}(x) (r_{n,1}(x, t) - \hat{r}_{n,1}(x, t)) \\ &+ \left((\hat{r}_{n,1}(x, t) - r_{n,1}(x, t)) + r_{n,1}(x, t) \right) (f_{n,1}(x) - \hat{f}_{n,1}(x)). \end{aligned}$$

Let $\beta_n = \sqrt{\frac{nh_n}{2\log(h_n^{-1})}}$. Because of (H.2), we have $\beta_n \rightarrow +\infty$ as $n \rightarrow +\infty$. Combining with the two previous corollaries and Equations (24), we see that:

$$\beta_n \sup_{t \in \mathbb{R}} \sup_{x \in I} |\text{Num}(26) - \widehat{g}_n(x, t)| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0 \quad (28)$$

where $\widehat{g}_n(x, t) = f_X(x)\mu_2(K)(\widehat{r}_{n,0}(x, t) - r_{n,0}(x, t)) - f_X(x)F(t|x)(\widehat{f}_{n,2}(x) - f_{n,2}(x))$.

Lemma 6.2. *The denominator of the expression (26), denoted $\text{Den}(26)$, satisfies:*

$$\sup_{x \in I} \left| \text{Den}(26) - \frac{1}{f_X(x)^2 \mu_2(K)} \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0. \quad (29)$$

Proof: For all $\epsilon > 0$, let the event

$$\mathcal{A}_\epsilon = \left\{ \sup_{x \in I} \left| \text{Den}(26) - \frac{1}{f_X(x)^2 \mu_2(K)} \right| > \epsilon \right\}$$

and for all $B > 0$, let the event

$$\mathcal{B} = \left\{ \sup_{x \in I} \left| (\widehat{f}_{n,2}(x)\widehat{f}_{n,0}(x) - \widehat{f}_{n,1}(x)^2) f_X(x)^2 \mu_2(K) \right| > B \right\}.$$

Then we have:

$$\mathbb{P}(\mathcal{A}_\epsilon) = \mathbb{P}(\mathcal{A}_\epsilon \cap \mathcal{B}) + \mathbb{P}(\mathcal{A}_\epsilon \cap \mathcal{B}^c).$$

In one hand, we have:

$$\mathbb{P}(\mathcal{A}_\epsilon \cap \mathcal{B}^c) \leq \mathbb{P}(\mathcal{B}^c) \leq \mathbb{P} \left(\sup_{x \in I} \left| \widehat{f}_{n,2}(x)\widehat{f}_{n,0}(x) - \widehat{f}_{n,1}(x)^2 \right| \leq \frac{B}{\inf_{x \in I} f_X(x)^2 \mu_2(K)} \right).$$

Taking $B = \frac{1}{2} \left(\inf_{x \in I} f_X(x)^2 \mu_2(K) \right)^2$, we obtain:

$$\mathbb{P}(\mathcal{A}_\epsilon \cap \mathcal{B}^c) \leq \mathbb{P} \left(\sup_{x \in I} \left| \widehat{f}_{n,2}(x)\widehat{f}_{n,0}(x) - \widehat{f}_{n,1}(x)^2 - f_X(x)^2 \mu_2(K) \right| \geq \frac{1}{2} \inf_{x \in I} f_X(x)^2 \mu_2(K) \right) \xrightarrow[n \rightarrow +\infty]{} 0.$$

This last limit is obtained by the following trivial decomposition:

$$\begin{aligned} & \widehat{f}_{n,2}(x)\widehat{f}_{n,0}(x) - \widehat{f}_{n,1}(x)^2 - f_X(x)^2 \mu_2(K) = \\ & \left((\widehat{f}_{n,2}(x) - f_{n,2}(x)) + (f_{n,2}(x) - f_X(x)\mu_2(K)) + f_X(x)\mu_2(K) \right) \\ & \times \left((\widehat{f}_{n,0}(x) - f_{n,0}(x)) + (f_{n,0}(x) - f_X(x)) \right) \\ & + \left((\widehat{f}_{n,2}(x) - f_{n,2}(x)) + (f_{n,2}(x) - f_X(x)\mu_2(K)) \right) f_X(x), \end{aligned}$$

and by applying the Corollary 6 and Equations (24), combined with the boundedness property of f_X on I .

On the other hand, we have:

$$\mathbb{P}(\mathcal{A}_\epsilon \cap \mathcal{B}) \leq \mathbb{P} \left(\sup_{x \in I} \left| \widehat{f}_{n,2}(x)\widehat{f}_{n,0}(x) - \widehat{f}_{n,1}(x)^2 - f_X(x)^2 \mu_2(K) \right| > \epsilon B \right) \xrightarrow[n \rightarrow +\infty]{} 0$$

for the same reason as before. We have then proved (29). \square

Combining (28) and (29), we have:

$$\beta_n \sup_{t \in \mathbb{R}} \sup_{x \in I} \left| (26) - \frac{\widehat{r}_{n,0}(x, t) - r_{n,0}(x, t)}{f_X(x)} - \frac{F(t|x)}{f_X(x)\mu_2(K)} \left(\widehat{f}_{n,2}(x) - f_{n,2}(x) \right) \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0. \quad (30)$$

This last limit is due to the following lemma.

Lemma 6.3. *Let $\mathcal{I} \subset \mathbb{R}^d$, with $d \in \mathbb{N}^*$, and X_n, Z_n, Y_n and Y random functions defined on \mathcal{I} such that*

$$\sup_{w \in \mathcal{I}} |X_n(w) - Z_n(w)| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0, \quad \sup_{w \in \mathcal{I}} |Y_n(w) - Y(w)| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0,$$

and for enough high $B > 0$, $\lim_{n \rightarrow +\infty} \mathbb{P}(\sup_{w \in \mathcal{I}} |X_n(w)| > B) = 0$ and Y is bounded on \mathcal{I} . Then,

$$\sup_{w \in \mathcal{I}} |X_n(w)Y_n(w) - Z_n(w)Y(w)| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0.$$

Proof:

$$\sup_{\mathcal{I}} |X_n Y_n - Z_n Y| \leq \sup_{\mathcal{I}} |X_n| \sup_{\mathcal{I}} |Y_n - Y| + \sup_{\mathcal{I}} |Y| \sup_{\mathcal{I}} |X_n - Z_n|,$$

then for all $\epsilon > 0$,

$$\mathbb{P} \left(\sup_{\mathcal{I}} |X_n Y_n - Z_n Y| > \epsilon \right) \leq \mathbb{P} \left(\sup_{\mathcal{I}} |X_n| \sup_{\mathcal{I}} |Y_n - Y| > \frac{\epsilon}{2} \right) + \mathbb{P} \left(\sup_{\mathcal{I}} |Y| \sup_{\mathcal{I}} |X_n - Z_n| > \frac{\epsilon}{2} \right).$$

Introducing the event $\mathcal{B} = \left\{ \sup_{\mathcal{I}} |Y| > B \right\}$ with $B > 0$, we bound the second term:

$$\mathbb{P} \left(\sup_{\mathcal{I}} |Y| \sup_{\mathcal{I}} |X_n - Z_n| > \frac{\epsilon}{2} \right) \leq \mathbb{P}(\mathcal{B}) + \mathbb{P} \left(\sup_{\mathcal{I}} |X_n - Z_n| > \frac{\epsilon}{2B} \right) \xrightarrow[n \rightarrow +\infty]{} 0$$

for enough high B . Now, we bound the first term:

$$\mathbb{P} \left(\sup_{\mathcal{I}} |X_n| \sup_{\mathcal{I}} |Y_n - Y| > \frac{\epsilon}{2} \right) \leq \mathbb{P} \left(\sup_{\mathcal{I}} |X_n| > B \right) + \mathbb{P} \left(\sup_{\mathcal{I}} |Y_n - Y| > \frac{\epsilon}{2B} \right) \xrightarrow[n \rightarrow +\infty]{} 0$$

for enough high B . \square

Let's studying now the second part (27) in the expression of $\widehat{F}_n^{(1)}(t|x) - \widehat{\mathbb{E}} \left(\widehat{F}_n^{(1)}(t|x) \right)$. Note that the numerator is equal to:

$$(r_{n,0}(x, t)f_{n,2}(x) - f_{n,1}(x)r_{n,1}(x, t)) \left(f_{n,2}(x)f_{n,0}(x) - f_{n,1}^2(x) - \widehat{f}_{n,2}(x)\widehat{f}_{n,0}(x) + \widehat{f}_{n,1}^2(x) \right).$$

The first term $\text{Num}_1(27) = r_{n,0}(x, t)f_{n,2}(x) - f_{n,1}(x)r_{n,1}(x, t)$ of this last expression converges uniformly, thanks to (24):

$$\sup_{t \in \mathbb{R}} \sup_{x \in I} |\text{Num}_1(27) - f_X(x)^2 \mu_2(K) F(t|x)| \xrightarrow[n \rightarrow +\infty]{} 0. \quad (31)$$

With the same arguments as in the study of the numerator of (26), we study the second term defined by:

$$\text{Num}_2(27) = f_{n,2}(x)f_{n,0}(x) - f_{n,1}^2(x) - \widehat{f}_{n,2}(x)\widehat{f}_{n,0}(x) + \widehat{f}_{n,1}^2(x)$$

and show that

$$\beta_n \sup_{x \in I} \left| \text{Num}_2(27) - f_X(x) \mu_2(K) \left(f_{n,0}(x) - \widehat{f}_{n,0}(x) \right) - f_X(x) \left(f_{n,2}(x) - \widehat{f}_{n,2}(x) \right) \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0. \quad (32)$$

Thanks to the boundedness property of f_X on I , and the Corollary 6, we have, by the Lemma 6.3:

$$\beta_n \sup_{t \in \mathbb{R}} \sup_{x \in I} \left| \text{Num}(27) - \widehat{j}_n(x, t) \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0$$

where $\widehat{j}_n(x, t) = f_X(x)^3 \mu_2(K) F(t|x) \left(\mu_2(K) \left(f_{n,0}(x) - \widehat{f}_{n,0}(x) \right) + \left(f_{n,2}(x) - \widehat{f}_{n,2}(x) \right) \right)$.

The denominator in (27) can be expressed as

$$\text{Den}(27) = \text{Den}(26) \frac{1}{f_{n,2}(x) f_{n,0}(x) - f_{n,1}(x)^2}.$$

It is clear that, thanks to (24)

$$\sup_{x \in I} \left| \frac{1}{f_{n,2}(x) f_{n,0}(x) - f_{n,1}(x)^2} - \frac{1}{f_X(x)^2 \mu_2(K)} \right| \xrightarrow[n \rightarrow +\infty]{} 0.$$

Then, thanks to the boundedness property of f_X on I , the Lemma 6.3 says that

$$\sup_{x \in I} \left| \text{Den}(27) - \frac{1}{f_X(x)^4 \mu_2(K)^2} \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0. \quad (33)$$

Finally, we have, thanks to the Lemma 6.3

$$\beta_n \sup_{t \in \mathbb{R}} \sup_{x \in I} \left| (27) - \frac{F(t|x)}{f_X(x)} \left(f_{n,0}(x) - \widehat{f}_{n,0}(x) \right) - \frac{F(t|x)}{f_X(x) \mu_2(K)} \left(f_{n,2}(x) - \widehat{f}_{n,2}(x) \right) \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0. \quad (34)$$

Remember that $\widehat{F}_n^{(1)}(t|x) - \widehat{\mathbb{E}} \left(\widehat{F}_n^{(1)}(t|x) \right) = (26) + (27)$. Then, combining (30) and (34) it is easy to see finally that

$$\beta_n \sup_{t \in \mathbb{R}} \sup_{x \in I} \left| \widehat{F}_n^{(1)}(t|x) - \widehat{\mathbb{E}} \left(\widehat{F}_n^{(1)}(t|x) \right) + \frac{F(t|x)}{f_X(x)} \left(\widehat{f}_{n,0}(x) - f_{n,0}(x) \right) - \frac{\widehat{r}_{n,0}(t, x) - r_{n,0}(t, x)}{f_X(x)} \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0. \quad (35)$$

Step 4:

Now choosing $c(x) = \frac{1}{f_X(x)}$ and $d(x) = -\frac{F(t|x)}{f_X(x)}$ in the definition of $W_{n,0}(x, t)$, the local empirical process, it is easy to show that:

$$\begin{aligned} W_{n,0}(x, t) &= \frac{nh}{f_X(x)} \widehat{r}_{n,0}(x, t) - \frac{F(t|x)}{f_X(x)} \times nh \widehat{f}_{n,0}(x) - \frac{nh}{f_X(x)} r_{n,0}(x, t) + \frac{F(t|x)}{f_X(x)} \times nh f_{n,0}(x) \\ &= \frac{nh}{f_X(x)} \left(\widehat{r}_{n,0}(x, t) - r_{n,0}(x, t) - F(t|x) \left(\widehat{f}_{n,0}(x) - f_{n,0}(x) \right) \right). \end{aligned}$$

Let $A_n = \sqrt{2nh_n \log(h_n^{-1})}^{-1}$, so we have:

$$A_n W_{n,0}(x, t) = \beta_n \left(\frac{\widehat{r}_{n,0}(t, x) - r_{n,0}(t, x)}{f_X(x)} - \frac{F(t|x)}{f_X(x)} \left(\widehat{f}_{n,0}(x) - f_{n,0}(x) \right) \right). \quad (36)$$

Then

$$\begin{aligned} \beta_n \left(\widehat{F}_n^{(1)}(t|x) - \widehat{\mathbb{E}} \left(\widehat{F}_n^{(1)}(t|x) \right) \right) &= A_n W_{n,0}(x, t) \\ &+ \beta_n \left(\widehat{F}_n^{(1)}(t|x) - \widehat{\mathbb{E}} \left(\widehat{F}_n^{(1)}(t|x) \right) + \frac{F(t|x)}{f_X(x)} \left(\widehat{f}_{n,0}(x) - f_{n,0}(x) \right) - \frac{\widehat{r}_{n,0}(t, x) - r_{n,0}(t, x)}{f_X(x)} \right), \end{aligned}$$

and applying Theorem 6.1

$$\sup_{t \in \mathbb{R}} \sup_{x \in I} \left| \beta_n \left(\widehat{F}_n^{(1)}(t|x) - \widehat{\mathbb{E}} \left(\widehat{F}_n^{(1)}(t|x) \right) \right) \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \sigma_F(I) \quad (37)$$

with

$$\begin{aligned} \sigma_F^2(I) &= \sup_{t \in \mathbb{R}} \sup_{x \in I} \mathbb{E} \left(\left[\frac{\mathbb{1}_{\{Y \leq t\}}}{f_X(x)} - \frac{F(t|x)}{f_X(x)} \right]^2 | X = x \right) f_X(x) \|K\|_2^2 \\ &= \sup_{t \in \mathbb{R}} \sup_{x \in I} \frac{\mathbb{E} \left([\mathbb{1}_{\{Y \leq t\}} - \mathbb{E}(Y \leq t | X = x)]^2 | X = x \right)}{f_X(x)} \|K\|_2^2 \\ &= \sup_{t \in \mathbb{R}} \sup_{x \in I} \left(\frac{F(t|x)(1 - F(t|x))}{f_X(x)} \right) \|K\|_2^2. \end{aligned}$$

This finishes the proof of Theorem 2.1. \square

6.2. Proof of Corollary 4. If the variable Y takes its values in $[\alpha, \beta]$, then

$$\widehat{m}_n^{(1)}(x) - m(x) = \int_{\alpha}^{\beta} y \left(\widehat{F}_n^{(1)}(dy, h_n|x) - F(dy|x) \right) = - \int_{\alpha}^{\beta} \left(\widehat{F}_n^{(1)}(y, h_n|x) - F(y|x) \right) dy.$$

It implies $\left| \widehat{m}_n^{(1)}(x) - m(x) \right| \leq |\beta - \alpha| \sup_{y \in \mathbb{R}} \left| \widehat{F}_n^{(1)}(y, h_n|x) - F(y|x) \right|$.

The conclusion of the Corollary 4 is then deduced from this last inequality combined with Corollary 3.

6.3. Proof of Corollary 5. Let $\alpha \in (0, 1)$. It suffices to show that

$$\mathbb{P} \left(\forall x \in I, \left| \widehat{q}_{\alpha,n}^{(1)}(x) - q_{\alpha}(x) \right| > \frac{2L_n(x)f_X(x)}{f_{X,Y}(x, q_{\alpha}(x))} \right) \longrightarrow 0 \text{ as } n \longrightarrow +\infty. \quad (38)$$

Let $\epsilon = \epsilon_{\alpha,x,n} = \frac{2L_n(x)f_X(x)}{f_{X,Y}(x, q_{\alpha}(x))}$. We have

$$\begin{aligned} \mathbb{P} \left(\forall x \in I, \left| \widehat{q}_{\alpha,n}^{(1)}(x) - q_{\alpha}(x) \right| > \epsilon \right) &= \mathbb{P} \left(\forall x \in I, \widehat{q}_{\alpha,n}^{(1)}(x) > q_{\alpha}(x) + \epsilon \right) + \mathbb{P} \left(\forall x \in I, \widehat{q}_{\alpha,n}^{(1)}(x) < q_{\alpha}(x) - \epsilon \right) \\ &= (I) + (II). \end{aligned}$$

To study the first term (I), consider the following implication, for all $x \in I$,

$$\left(\widehat{q}_{\alpha,n}^{(1)}(x) = \inf \{ t \in \mathbb{R} : \widehat{F}_n^{(1)}(t, h_n|x) \geq \alpha \} > q_{\alpha}(x) + \epsilon \right) \Rightarrow \left(\widehat{F}_n^{(1)}(q_{\alpha}(x) + \epsilon, h_n|x) < \alpha \right).$$

Then

$$\begin{aligned} (I) &\leq \mathbb{P} \left(\forall x \in I, \widehat{F}_n^{(1)}(q_{\alpha}(x) + \epsilon, h_n|x) < \alpha \right) \\ &= \mathbb{P} \left(\forall x \in I, \widehat{F}_n^{(1)}(q_{\alpha}(x) + \epsilon, h_n|x) - F(q_{\alpha}(x) + \epsilon|x) < \alpha - F(q_{\alpha}(x) + \epsilon|x) \right) \\ &\leq \mathbb{P} \left(\forall x \in I, \sup_{y \in \mathbb{R}} \left| \widehat{F}_n^{(1)}(y, h_n|x) - F(y|x) \right| > F(q_{\alpha}(x) + \epsilon|x) - \alpha \right) \\ &\leq \mathbb{P} \left(\forall x \in I, \sup_{y \in \mathbb{R}} \left| \widehat{F}_n^{(1)}(y, h_n|x) - F(y|x) \right| > \frac{\epsilon f_{X,Y}(x, q_{\alpha}(x))}{2f_X(x)} \right) \\ &= \mathbb{P} \left(\forall x \in I, \sup_{y \in \mathbb{R}} \left| \widehat{F}_n^{(1)}(y, h_n|x) - F(y|x) \right| > L_n(x) \right) \longrightarrow 0 \text{ as } n \longrightarrow +\infty \end{aligned}$$

where $F(q_{\alpha}(x) + \epsilon|x) - \alpha = F(q_{\alpha}(x) + \epsilon|x) - F(q_{\alpha}(x)|x) = \epsilon \frac{f_{X,Y}(x, q_{\alpha}(x))}{f_X(x)} + o(\epsilon) \geq \frac{\epsilon f_{X,Y}(x, q_{\alpha}(x))}{2f_X(x)}$.

To study the second term (II), consider the following implications, for all $x \in I$:

$$\begin{aligned} \left(\hat{q}_{\alpha,n}^{(1)}(x) < q_\alpha(x) - \epsilon \right) &\Rightarrow \left(F\left(\hat{q}_{\alpha,n}^{(1)}(x)|x\right) < F(q_\alpha(x) - \epsilon|x) < \alpha \leq \hat{F}_n^{(1)}\left(\hat{q}_{\alpha,n}^{(1)}(x), h_n|x\right) \right) \\ &\quad \text{by the growing property of } F \text{ and the definition of the quantile } \hat{q}_{\alpha,n}^{(1)}(x) \\ &\Rightarrow \left(\hat{F}_n^{(1)}\left(\hat{q}_{\alpha,n}^{(1)}(x), h_n|x\right) - F\left(\hat{q}_{\alpha,n}^{(1)}(x)|x\right) \geq \alpha - F(q_\alpha(x) - \epsilon|x) \right). \end{aligned}$$

Then

$$\begin{aligned} (II) &\leq \mathbb{P}\left(\forall x \in I, \hat{F}_n^{(1)}\left(\hat{q}_{\alpha,n}^{(1)}(x), h_n|x\right) - F\left(\hat{q}_{\alpha,n}^{(1)}(x)|x\right) \geq \alpha - F(q_\alpha(x) - \epsilon|x)\right) \\ &\leq \mathbb{P}\left(\forall x \in I, \sup_{y \in \mathbb{R}} \left| \hat{F}_n^{(1)}(y, h_n|x) - F(y|x) \right| > \alpha - F(q_\alpha(x) - \epsilon|x)\right) \\ &\leq \mathbb{P}\left(\forall x \in I, \sup_{y \in \mathbb{R}} \left| \hat{F}_n^{(1)}(y, h_n|x) - F(y|x) \right| > \frac{\epsilon f_{X,Y}(x, q_\alpha(x))}{2f_X(x)}\right) \\ &= \mathbb{P}\left(\forall x \in I, \sup_{y \in \mathbb{R}} \left| \hat{F}_n^{(1)}(y, h_n|x) - F(y|x) \right| > L_n(x)\right) \longrightarrow 0 \text{ as } n \longrightarrow +\infty \end{aligned}$$

where $\alpha - F(q_\alpha(x) - \epsilon|x) = F(q_\alpha(x)|x) - F(q_\alpha(x) - \epsilon|x) = \epsilon \frac{f_{X,Y}(x, q_\alpha(x))}{f_X(x)} + o(\epsilon) \geq \frac{\epsilon f_{X,Y}(x, q_\alpha(x))}{2f_X(x)}$.

Finally, we have proved (38).

ACKNOWLEDGEMENT

We thank to Doctors Jean-Pierre Masutti and Alain Miton of the Service de foetopathologie et de placentologie of the Maternité Régionale Universitaire (CHU Nancy, France), for the fetal data.

REFERENCES

- [1] D.G. Altman, L.S. Chitty, *Charts of fetal size: 1. Methodology.*, Br. J. Obstet. Gynaecol., 101 (1994), pp. 29-34.
- [2] A. Berlinet, L. Devroye, *A comparison of kernel density estimates*, Publ. Inst. Statist. Univ. Paris, 38 (1994), pp. 3-79.
- [3] D. Blondin, *Lois limites uniformes et estimation non paramétrique de la régression*, PhD thesis, Université de Paris 6, France, 2004.
- [4] G. Collomb, *Estimation non paramétrique de probabilités conditionnelles*, C. R. Acad. Sci. Paris Sér. A-B, 291 (1980), pp. 427-430.
- [5] G. Collomb, *Estimation non-paramétrique de la régression: revue bibliographique*, Int. Statist. Rev., 49 (1981), pp. 75-93.
- [6] P. Deheuvels, *Estimation non-paramétrique de la densité par histogramme généralisé*, La Revue de Statistique Appliquée, 35 (1977), pp. 5-42.
- [7] P. Deheuvels, *Limit laws for kernel density estimators for kernels with unbounded supports*, In Asymptotics in Statistics and Probability, (Ed., M. L. Puri), V.S.P., Amsterdam, 2000.
- [8] P. Deheuvels, G. Derzko, *Asymptotic certainty bands for kernel density estimators based upon a bootstrap resampling scheme*, Statistical Models and Methods for Biomedical and Technical Systems, 3 (2008), pp. 171-186.

- [9] P. Deheuvels, D.M. Mason, *General asymptotic confidence bands based on kernel-type function estimators*, Statist. Infer. Stochastic Process, 7(3) (2004), pp. 225-277.
- [10] L. Devroye, *The uniform convergence of the Nadaraya-Watson regression function estimate*, Can. J. Statist., 6 (1978), pp. 179-191.
- [11] J. Dony, U. Einmahl, D.M. Mason, *Uniform in Bandwidth Consistency of Local Polynomial Regression Function Estimators*, Austrian Journal of Statistics, 35(2-3) (2006), pp.105-120.
- [12] U. Einmahl, D.M. Mason, *An empirical process approach to the uniform consistency of kernel-type function estimators*, J. Theoret. Probab., 13(1) (2000), pp. 1-37.
- [13] U. Einmahl, D.M. Mason, *Uniform in bandwidth consistency of kernel-type function estimators*, Ann. Statist., 33(3) (2005), pp. 1380-1403.
- [14] J. Fan, I. Gijbels, *Local polynomial modeling and its applications*. Monographs on Statistics and Applied Probability, Chapman and Hall, Vol. 66, 1996.
- [15] S. Ferrigno, *Un test d'adéquation global pour la fonction de répartition conditionnelle*, PhD thesis, Université de Montpellier 2, France, 2004.
- [16] A. Gannoun, S. Girard, C. Guinot, J. Saracco, *References curves based on nonparametric quantile regression*, Statistics in Medicine, 21 (2002), pp. 3119-3155.
- [17] A.M. Guihard-Costa, *Les variations des vitesses de croissance au cours de la vie foetale*, Bulletins et Mémoires de la Société d'anthropologie de Paris, Nouvelle Série, 5(1-2) (1993), pp.11-20.
- [18] W. Härdle, *Applied Nonparametric Regression*, Cambridge University Press, Cambridge, 1990.
- [19] W. Härdle, P. Janssen, R. Serfling, *Strong uniform consistency rates for estimators of conditional functionals*, Ann. Statist., 16(4) (1988), pp.1428-1449.
- [20] Q. Li, J. Lin, J. S. Racine, *Optimal Bandwidth Selection for Nonparametric Conditional Distribution and Quantile Functions*, Journal of Business and Economic Statistics, 31(1) (2013), pp.57-65.
- [21] M. Mint El Mouvid, *Sur l'estimateur linéaire local de la fonction de répartition conditionnelle*, PhD thesis, Université de Montpellier 2, France, 2000.
- [22] E.A. Nadaraya, *On estimating regression*, Theor. Probab. Appl., 9 (1964), pp. 141-142.
- [23] E. Parzen, *On estimation of a probability density function and mode*, Ann. Math. Statist., 33 (1962), pp. 1065-1076.
- [24] M. Rosenblatt, *Remarks on some nonparametric estimates of a density function*, Ann. Math. Statist., 27 (1956), pp. 832-837.
- [25] P. Royston, E.M. Wright, *How to construct "normal ranges" for fetal variables*, Ultrasound Obstet. Gynecol., 11 (1998), pp. 30-8.
- [26] W. Stute, *A law of the iterated logarithm for kernel density estimators*, Ann. Probab., 10(2) (1982), pp. 414-422.

- [27] W. Stute, *On almost sure convergence of conditional empirical distribution functions*, Ann. Probab., 14(3) (1986), pp. 891-901.
- [28] M. Talagrand, *Sharper bounds for Gaussian and empirical processes*, Ann. Probab., 22 (1994), pp. 28-76.
- [29] A.B. Tsybakov, *Introduction to Nonparametric Estimation*, Springer Series in Statistics, 2008.
- [30] G.S. Watson, *Smooth regression analysis*. Sankhyā Ser.A, 26 (1964), pp. 359-372.
- [31] E. Youndje, *Convergence properties of the kernel estimator of conditional density*, Rev. Roumaine Math. Pures Appl., 41(7-8), 1996.
- [32] E. Youndje, *Contribution à l'estimation non-paramétrique par la méthode du noyau*, Habilitation à diriger des recherches, Université de Rouen et du Havre, France, 2011.